ABSTRACT

A Multi-level Analysis of the Spread of COVID-19

Bennet E. Hickok, M.S.Eco

Chairperson: James West, Ph.D.

This paper uses extensions of the traditional methods for evaluating panel data to evaluate the effect of Non-Pharmaceutical Interventions (NPI) on the spread of COVID-19. I utilize data from weather conditions, policy interventions, past outcomes, and political landscapes at the county level. These components allow me to navigate confounding issues with traditional models such as heterogeneity, endogeneity, and measurement error. The results of this model support the efficacy of policy interventions. I also find that poor weather conditions contribute to the spread of the disease, which indicates that the disease spreads less effectively outdoors. Finally, I find that the share of GOP voters in the previous election is positively associated with the spread of the disease. The ability to combine time variant and invariant components with minimal assumption, makes this model a helpful foundation for further research.

A Multi-level Analysis of the Spread of COVID-19

by

Bennet E. Hickok, B.S.

A Thesis

Approved by the Department of Economics

Charles M. North, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Master of Science in Economics

Approved by the Thesis Committee

James West, Ph.D., Chairperson

Michael Richards, M.D. Ph.D.

Jonathan Trower, Ph.D.

Accepted by the Graduate School
May 2021

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF TABLES ... vi

ACKNOWLEDGMENTS ... vii

DEDICATION ... viii

CHAPTER ONE
Introduction ... 1

CHAPTER TWO
Methodology ... 4

CHAPTER THREE
Data and Variables ... 6
Cumulative Cases ... 6
Policy ... 6
Information ... 7
Weather Conditions ... 7
Fixed Effects ... 8

CHAPTER FOUR
Results ... 10

CHAPTER FIVE
Discussion ... 14

CHAPTER SIX
Conclusion ... 17

APPENDIX A
Alternative Dependent Variables ... 20

APPENDIX B
Regression Model Comparison ... 21

APPENDIX C
Auxiliary Regression of Policy and Behavior ... 22

APPENDIX D
Distributed Lags ... 23

APPENDIX E
Bucketed Regressions 24

BIBLIOGRAPHY 25

LIST OF TABLES

Table 3.1. Summary Statistics 9

Table 4.1. Comparison of Policies 11

Table 4.2. Summary of Results 13

Table 5.1. Summary of Missing Mobility Data 15

Table A.1. Alternative Dependent Variables 20

Table B.1. Comparison of Regression Models 21

Table D.1. Comparison of Distributed Lags 23

Table E.1. Table of Regressions on Stratified Data 24

ACKNOWLEDGMENTS

I would like to thank Dr. West for his mentorship throughout this process. It is a daunting task to study something that is unfolding while it is being studied and Dr. West's guidance has been invaluable. I would also like to thank Dr. Pham who made this process possible. My journey to this point has not been a linear one and Dr. Pham has helped me navigate the uncertain waters.

To my wife, Katie
She has supported me through everything, and she always pushes me to be better. I hope this work reflects that support. I am forever grateful.

CHAPTER ONE

Introduction

In 2019 a new kind of disease passed from bats to humans leading to a pandemic of proportions not seen since 1918. This disease was unique in its ability to spread before symptoms emerged with some carriers spreading the disease without ever showing symptoms. This feature meant that standard practices of quarantining and contact tracing were less effective than they would typically be (7).

It also posed a unique challenge for epidemiologists who have spent centuries perfecting precise models that could tell exactly how a disease would spread and what interventions would be most effective. These models are remarkable in their simplicity and accuracy but their validity is dependent on the promptness and accuracy of the input data. With COVID-19, there were a number of unknowns that standard models were not designed to account for. This has led to the unification of epidemiologists and other social scientists in the fight against this disease (8).

Much of the literature that has been published around COVID-19 has come from economists who are no strangers to uncertainty or unknowns. The most compelling example of this interaction has been the body of work that has emerged around the SIR-model. The canonical SIR-model splits the population into three buckets: Susceptible, Infected, and Recovered. The Recovered are assumed to be impervious to the disease which makes the main question how quickly the Infected group spreads the disease to the Susceptible population. Different diseases have baseline rates of contagion known as r_0.

If r_0 is less than 1, then infected individuals will recover more quickly than others will be infected and the disease will die out on its own.

If r_0 is greater than 1, then the disease will spread if left unattended. At this point, the goal of policy-makers would be to reduce the level of contagion at any given time (r_t) to a level below 1. This will occur naturally as the Susceptible individuals get infected and move into the Recovered bucket. This phenomenon is what leads to the bell-curve shape that people refer to when they discuss ways to "flatten the curve". Alternatively, it can be done by reducing human contact through stay-at-home orders, reducing the probability of people getting sick if they are exposed by encouraging face coverings, or mechanically removing people from the Susceptible bucket with vaccines.

In practice, this model quickly breaks down if you do not have a clear picture of how many people are Susceptible, Infected, or Recovered. This is where economists have focused much of their attention. There is a large body of work that has added "exposed" or "asymptomatic" buckets to the canonical SIR model in hopes that these SEIR or SAIR models will fit better to the data that are emerging. Others have used Bayesian analysis and Monte Carlo sampling to account for the heterogeneity that has been found when calibrating the SIR model to the data (2).

Another important issue that hinders the ability to make effective policy recommendations is endogeneity. There are two possible endogenous processes that cloud the identification of the effect of policy changes. The first is the impact of infection rates on government policies because while policies can impact infection rates, they are also made in response to infection rates. The second endogenous process is the interplay between infection rates and activity. If there is a disease outbreak in my community, it is reasonable to assume that I would stay home where

These endogenous processes would bias the effects of policies and community efforts to limit exposure to the disease because locations with the greatest responses are likely to be the locations most impacted by the disease. In fact, this can lead to the uncomfortable scenario where the data show that aggressive policies are associated with a greater number of COVID-19 cases. It can also yield the equally unsettling result that locations with greater levels of activity end up with fewer instances of COVID-19 even though it would be logical to assume that in fact people are more active because the disease is less pervasive in their community (6).

This paper will seek to enter into the discussion by evaluating the efficacy of policy decisions, building on the body of literature mentioned above to resolve issues of endogeneity, simultaneity and heterogeneity. Additionally, this paper addresses some key data concerns with the existing body of work to corroborate findings where possible and determine where further consideration is needed.

CHAPTER TWO

Methodology

I follow the empirical approaches that Karaivanov et. al. (2020) used in Canada with some important modifications that will be discussed in subsequent sections. In order to assess the impact of non-pharmaceutical interventions (NPI) on the prevalence of COVID-19, this paper leverages the following model:

$$Y_{it} = \beta_0 + \beta_1 P_{it-l} + \beta_2 I_{it-l} + \beta_3 W_{it-l} + \beta_4 F_i + \epsilon_{it}$$

Where:

i =The specific county in the cross section.

t =The time designated in days.

l =The time lag, measured in days.

This equation models the relationship between the number of new COVID-19 (Y_{it}) cases and the previous policy measures (P_{it-l}), the lagged number of new cases (Y_{it-l}), the weather conditions(W_{it-l}) and the fixed effects impacting the various cross-sectional units (F_i).

In order to evaluate the number of new COVID-19 cases, I take the 7-day moving average of the number of new cases per 100,000 people. This is a significant departure from much of the early literature, which used the growth rate as the dependent variable.

While the exponential growth rate is very helpful for fitting the data to existing models and establishing counterfactuals for what could have been under different scenarios, it has some serious limitations. Namely, it tends to fit epidemic models very well early on when the spread is clearly exponential. However, as the spread of the disease subsides, an exponential growth model does not fit the data well. That is especially so with the pandemic at hand where re-openings have led to multiple peaks and important deviations from the standard "curve" (4).

It is difficult to ensure that all cases are reported instantaneously and patterns such as batch reporting on a certain day could give the appearance of time trends that don't exist. As a result, I use the 7-day moving average to smooth these outliers and ensure that the true signals are clear.

Additionally, I follow the lead of Chernozhukov et. al. (2020) in using a 14 day lag when observing the determinants of the COVID-19 case rate. This accounts for the time it takes to manifest symptoms (incubation period), the time it takes to report symptoms once they are observed (reporting period) and any lag that may exist between a change in a determinant and the implementation of that change (e.g. a policy change could take a few days to take effect or information changes may not percolate through the community instantaneously.)

I use a Hybrid model with robust standard errors. This allows me to conduct the most granular analysis possible under the weakest assumptions available to us. I maintain the unbiased estimates that I would expect under a Fixed Effects model, while gaining additional insights into the impact that time-invariant factors have on the dependent variables. In the appendix, I include similar analysis done

with pure Fixed Effects and Random Effects regressions as well as a Correlated Random Effects model (Schunck 2013).

CHAPTER THREE

Data and Variables

Cumulative Cases

Most studies that have been done on this topic have leveraged data from the New York Times or from Johns Hopkins to estimate the prevalence of COVID-19 cases. I began our study using both data sources in parallel with the hope that the differences between the results would be immaterial.

This was the case when I reviewed the data at the county level or for the largest counties. However, for smaller counties, I saw strange phenomena, where, for example, Johns Hopkins was reporting thousands of cases for counties with less than a thousand residents. Typically, these were counties with large hospitals or universities. That raises the interesting question of how to account for a college student from Minnesota who contracts COVID-19 in Texas but is in Minnesota for much of the evaluation period due to the move to remote learning or other patients from undefined areas who are diagnosed and treated at a large regional hospital. Ultimately, the New York Times dataset was more closely correlated with the population data reported from other sources. Accordingly, I used these data alone, which led to the fewest outliers (12).

Policy

In order to study the impact of various policies and NPIs, I utilized the state-level data from the Oxford COVID-19 Government Response Tracker(5). These data included daily indexes of the overall response that could then be decomposed into

- *Containment Health (C)*- Includes such items as school closings, workplace closing, and stay-at-home orders.

- *Economic (E)*- Includes stimulus, subsidies, and other economic policies to help individuals and business.
- *Health (H)*- Evaluates the availability of hospital beds, tests, investment in vaccines, and the use of face coverings.

These categories are then available as indices, which combine these elements in various combinations1.

Information

The information variable in our model is the lagged outcome variable. I include this to account for autocorrelation wherein having more cases in a given period is likely to result in a larger number of cases in subsequent periods. This inclusion allows me to account for elements of the growth curve without explicitly including growth rates in the model.

Weather Conditions

I determined the weather conditions by evaluating the amount that the average temperature on a given day deviated from a midpoint of 70 degrees. I gathered these data

from the National Oceanic and Atmospheric Administration (9). I considered this variance in absolute terms under the assumption that people would stay inside if it was either too cold or too hot. I then took the square root of that absolute deviation. The intuition here is that the incremental effect of an additional degree will decrease as you get further from the midpoint. Some people might view 60 degrees as too cold and another group might draw the line at 55 degrees. But, once you get down to freezing, 5 more degrees would have less of an impact on a person's propensity to go outside.

Fixed Effects

In my analysis, I have used the election results from the 2020 election to determine whether a county voted primarily democrat or republican (3). For my main analysis, I have used the raw percentages although I have included additional analysis in Appendix E with the results bucketed into a categorical variable. It is well established that other factors such as race and income are related to the spread of the disease but it is less clear whether those factors impact the effectiveness of various interventions.

However, there is evidence that Fox News viewership has some association with non-compliance to COVID-19 policies and as such would be an important factor in the efficacy of those policies (11). In their analysis, Simonov et al. find that due to changes in methodology, there is a tradeoff between promptness and completeness in the data. Thus, I use election results directly to proximate their analysis with similar results.

Summary statistics for the variables described in this section can be found in the following table.

Table 3.1. Summary Statistics

VARIABLES	(1) mean (sd)
NewCaseRate	16.98
	(25.42)
GrowthRate	0.0146
	(0.0906)
GrowthOfGrowthRate	-0.000823
	(0.0749)
DurationWeightedNewCases	0.151
	(0.290)
weather	3.155
	(1.410)
GovernmentResponseIndex	49.44
	(9.670)
ContainmentHealthIndex	50.60
	(9.226)
EconomicSupportIndex	41.89
	(23.22)
StringencyIndex	52.05
	(14.03)
per_gop	61.66
	(15.80)
Observations	509,184

CHAPTER FOUR

Results

My study finds that Non-Pharmaceutical Interventions are effective in reducing the spread of COVID-19. Even when controlling for endogeneity that may arise from governments and communities responding to previous levels of contagion, I found that higher levels of stringency around economic policy, healthcare infrastructure, and stay at home orders, all reduce the number of new COVID-19 cases. Additionally, the weather conditions have an important impact on the spread of COVID-19, with extreme temperatures leading to a larger number of new cases. Finally, the percentage of voters who voted republican in the 2020 elections in a given county had a positive association with the spread of COVID-19 in that county.

During the period from March 2020 to December 2020, the average US County saw a baseline number of 18 new COVID-19 cases per day for every 100,000 people. Although pandemics rarely follow a linear path due to their compounding nature, the series of openings, closings, and re-openings that took place in the US gave a path that better approximates a linear trend than I would typically expect.

As measured by the OxCGRT Government Response Index, a 10% increase in stringency would lead to 4 fewer cases per 100,000. Based on this, registering 100% on the policy index could lead to 20 more people recovering each day than becoming sick. Disaggregating the stringency index into its component can help give an idea of which policies are having the greatest impact.

From table 4.1, I see that the government response index (which includes elements from each category) has the greatest impact. Implementing economic support policies alone is approximately 31% as effective. With all else being equal, a county

Table 4.1. Comparison of Policies

VARIABLES	(1) NewCaseRate	(2) NewCaseRate	(3) NewCaseRate	(4) NewCaseRate
L14.dNewCaseRate	0.933***	0.937***	0.951***	0.894***
	(0.00155)	(0.00155)	(0.00155)	(0.00158)
L14.dweather	1.378***	1.395***	1.278***	1.844***
	(0.0244)	(0.0244)	(0.0246)	(0.0244)
L14.dGovernmentResponseIndex	-0.392***			
	(0.00474)			
per_gop	0.0521***	0.0523***	0.0493***	0.0524***
	(0.00451)	(0.00444)	(0.00460)	(0.00425)
mNewCaseRate	1.178***	1.188***	1.166***	1.188***
	(0.00850)	(0.00864)	(0.00768)	(0.00827)
mweather	1.320***	1.325***	1.559***	1.305***
	(0.162)	(0.161)	(0.163)	(0.155)
mGovernmentResponseIndex	0.0178*			
	(0.0106)			
L14.ContainmentHealthIndex		-0.369***		
		(0.00477)		
mContainmentHealthIndex		0.413***		
		(0.0123)		
L14.dEconomicSupportIndex			-0.129***	
			(0.00228)	
mEconomicSupportIndex			-0.0140***	
			(0.00388)	
L14.dStringencyIndex				-0.337***
				(0.00273)
mStringencyIndex				0.0377***
				(0.00867)
Constant	-7.021***	-8.563***	-5.789***	-8.211***
	(0.864)	(0.897)	(0.645)	(0.796)
Observations	290,367	290,367	291,029	290,363
Number of fips	2,762	2,762	2,762	2,762

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

that implements economic support policies alone would see 6 new COVID-19 cases per day.

Another factor associated with the number of new cases is the political makeup of a county. Based on this model, an average county that leans aggressively republican would be expected to have 21 new cases per day while a similar county leaning democratic would have 16 new cases per day. It is likely that the mechanism for this association is the messaging and subsequent policy compliance within each party.

I included weather to help better understand the impact of activity. Having more pleasant weather does not mean that people will be less active-in fact our auxiliary regression shows that it makes them more active. However, when the conditions are amenable, people are more likely to spend time outdoors where COVID-19 spreads less effectively. This is consistent with what I see in the data with an expected baseline number of 19 new cases following a 70-degree day and 32 new cases following a day with extreme temperatures. The full table of results can be found in table 4.2.

Table 4.2. Summary of Results

	(1)
VARIABLES	NewCaseRate
L14.dNewCaseRate	0.933***
	(0.00155)
L14.dweather	1.378***
	(0.0244)
L14.dGovernmentResponseIndex	-0.392***
	(0.00474)
per_gop	0.0521***
	(0.00451)
mNewCaseRate	1.178***
	(0.00850)
mweather	1.320***
	(0.162)
mGovernmentResponseIndex	0.0178*
	(0.0106)
Constant	-7.021***
	(0.864)
Observations	290,367
Number of fips	2,762

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

CHAPTER FIVE

Discussion

Only needed for traditional dissertation format. The main contribution of this paper is not in the novelty of the results but in the integrity of the data. Many of my findings corroborate the consensus that has emerged in the literature thus far but the methodologies depart in some important ways. These departures make this model a potential baseline for future research.

The first departure was the use of the Hybrid Random effects model. This model allows me to estimate the impact of both level 1 and level 2 variables without the bias that can emerge with a traditional Random Effects models if the underlying assumptions are not met(10). This is true of both Hybrid models and Correlated Random Effects models and in fact I get similar results under both models. However, Hybrid models perform better when random slopes are introduced. Although random slopes are not included in this paper, our initial investigation shows that the effects of various treatments could vary significantly across panels making this a very helpful feature for further research.

Another important departure is the omission of the Behavior component in the model. Most papers that have come before this one have used the Google Mobility data to estimate the behavioral changes throughout the Pandemic(Goo). This had the distinct advantage of breaking activity down into categories such as Retail, Work, Groceries, and Parks. However, there were some important gaps in the data present as shown in table 5.1.

In addition to the magnitude of these omissions, the type of omission is also concerning. Google intentionally omitted samples under a certain number to protect the identities of individuals within those samples. Moreover, they only have access

Table 5.1. Summary of Missing Mobility Data

Variable	Missing	Total	Percent Missing
Retail	225,640	612,060	36.87
Grocery	267,084	612,060	43.64
Parks	474,356	612,060	77.50
Transit	382,159	612,060	62.44
Workplaces	14,776	612,060	2.41
Residential	282,890	612,060	46.22

to users of Android devices. Given the difference in price points between Android, Apple, and other devices, any factors that impact the price elasticity for the demand of phones (such as income, age, or occupation) could further bias the sample. Based on this, it is unlikely that these records are Missing Completely At Random (MCAR) or even Missing At Random (MAR) which subverts any meaningful attempt to impute the missing records. This also means that any regression based on the truncated sample is less reliable.

Putting all this aside, there are some important issues with the inclusion of behavior itself. If I am to believe that policy makers respond to signals based on the severity of the pandemic at any given time, it would stand to reason that the general population would respond to those same signals. Thus, a model that includes the lagged severity of the Pandemic, the Policy response, and the behavioral response; would effectively be capturing the same portion of the variance in the data with three separate variables. This would undermine the validity of all of those estimates.

Moreover, not all activity is the same. Walking your dog around the block will not put you in contact with as many people as would going to a dog park. Similarly attending an outdoor event where you practice social distancing would not be as risky as packing into a crowded arena or other building. Thus, even the more granular categories that

Google provides do not capture all the relevant information needed to determine the impact of behavior.

I sidestep these issues entirely by replacing the behavioral component with a measure of the weather conditions. This is effectively an instrument for the type of activity under the assumption that you are more likely to choose an outdoor event over the crowded arena when weather conditions are more pleasant. Additionally, it prevents the redundancy mentioned above that you would get by accounting for endogeneity multiple times.

The variants included in this section are beyond the scope of this paper. However, I mention them because I think they highlight the importance of this model as a foundation for further research. By identifying a model that is flexible and robust under the weakest possible assumptions, and then combining that with complete and timely data sources, this paper serves as a helpful starting point for answering any number of questions around how to properly respond to a pandemic.

CHAPTER SIX

Conclusion

COVID-19 has had an indelible effect on the world. It spread across every continent, changing the way people work, trade, and interact with each other. Unlike diseases that have come before, COVID-19 is contagious before any symptoms emerge and it manifests in dramatically different ways for different hosts. The unknowns inherent to COVID-19 have challenged traditional models for tracking and preventing the spread of the disease. In the face of these variables, economists have brought a unique perspective to the spread of the disease.

In this paper, I have built on extensions of traditional panel data methodologies to incorporate time variant and invariant factors in an analysis of the spread of COVID-19. This has allowed me to explore sources of endogeneity and heterogeneity under the weakest possible assumptions. I have also sought to choose data sources for this model that address possible sources of error. I have sidestepped the use of mobility data in order to ensure full coverage of our data. Additionally, I have chosen to use data from the New York Times to address irregularities found in other sources.

The results of this model corroborate much of the work that has been done on the subject to this point. I find that COVID-19 spreads most aggressively under poor weather conditions, which indicates that the disease spreads most effectively indoors. I also find, that policy interventions are effective across the board. Economic support has a marginal impact on its own but has a much stronger impact when combined with other interventions. Meanwhile, lock-downs and quarantines have a substantial impact on their own but these too are more effective when combined with other policies. While each of the policies are effective individually, the interventions are most effective when implemented in tandem.

In addition to validating other research that is coming out on this topic, this paper creates a solid foundation for further research. The combination of weak assumptions and robust data sources allows for several variations and extensions to understand the nuances of the spread of COVID-19. It is likely that COVID-19 will not be the last pandemic this world will see so the more we can learn the better.

APPENDICES

APPENDIX A

Alternative Dependent Variables

Table A.1. Alternative Dependent Variables

VARIABLES	(1) New Case Rate	(2) Growth Rate	(3) Growth of Growth Rate	(4) Duration Weighted New Cases
L14.dNewCaseRate	0.933***			
	(0.00155)			
L14.dweather	1.378***	0.000825***	-0.000147	0.0240***
	(0.0244)	(0.000160)	(0.000120)	(0.000346)
L14.dGovernmentResponseIndex	-0.392***	-0.00126***	5.15e-05**	-0.00387***
	(0.00474)	(3.15e-05)	(2.37e-05)	(6.68e-05)
per_gop	0.0521***	4.65e-05	2.04e-06	-0.000247
	(0.00451)	(3.93e-05)	(1.22e-05)	(0.000246)
mNewCaseRate	1.178***			
	(0.00850)			
mweather	1.320***	0.000469	0.00107**	0.0221**
	(0.162)	(0.00139)	(0.000447)	(0.00864)
mGovernmentResponseIndex	0.0178*	0.000203**	-5.00e-06	-0.000886*
	(0.0106)	(8.23e-05)	(2.57e-05)	(0.000524)
L14.dGrowthRate		0.0130***		
		(0.00207)		
mGrowthRate		0.953***		
		(0.0592)		
L14.dGrowthOfGrowthRate			0.141***	
			(0.00218)	
mGrowthOfGrowthRate			1.057***	
			(0.0341)	
L14.dDurationWeightedNewCases				0.260***
				(0.00151)
mDurationWeightedNewCases				0.850***
				(0.0117)
Constant	-7.021***	-0.0145**	-0.00391*	0.00575
	(0.864)	(0.00709)	(0.00221)	(0.0440)
Observations	290,367	258,398	230,049	290,367
Number of fips	2,762	2,742	2,727	2,762

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX B

Regression Model Comparison

Table B.1. Comparison of Regression Models

VARIABLES	(1) Hybrid	(2) Correlated Random Effects	(3) Random Effects	(4) Fixed Effects
L14.dNewCaseRate	0.933***			
	(0.00155)			
L14.dweather	1.378***			
	(0.0244)			
L14.dGovernmentResponseIndex	-0.392***			
	(0.00474)			
per_gop	0.0521***	0.0521***	0.0278***	
	(0.00451)	(0.00451)	(0.00580)	
mNewCaseRate	1.178***	0.245***		
	(0.00850)	(0.00862)		
mweather	1.320***	-0.0584		
	(0.162)	(0.164)		
mGovernmentResponseIndex	0.0178*	0.410***		
	(0.0106)	(0.0116)		
L14.NewCaseRate		0.933***	0.939***	0.937***
		(0.00155)	(0.00153)	(0.00155)
L14.weather		1.378***	1.371***	1.371***
		(0.0244)	(0.0241)	(0.0243)
L14.GovernmentResponseIndex		-0.392***	-0.358***	-0.390***
		(0.00474)	(0.00445)	(0.00472)
Constant	-7.021***	-7.021***	16.98***	20.24***
	(0.864)	(0.864)	(0.460)	(0.247)
Observations	290,367	290,367	292,596	292,596
R^2				0.603
Number of fips	2,762	2,762	2,763	2,763

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX C

Auxiliary Regression of Policy and Behavior

	(1)
VARIABLES	activity
StringencyIndex	0.000112***
weather	0.00751***
Constant	0.192***
	(0.000166)
Observations	598,287
Number of fips	2,797
R^2	0.150

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX D

Distributed Lags

Table D.1. Comparison of Distributed Lags

VARIABLES	(1) NewCaseRate
L14.NewCaseRate	0.729***
	(0.00371)
L7.weather	2.235***
	(0.0787)
L14.weather	0.0780
	(0.0833)
L21.weather	0.154**
	(0.0634)
L28.weather	1.181***
	(0.0853)
L7.GovernmentResponseIndex	0.0568*
	(0.0309)
L14.GovernmentResponseIndex	-0.407***
	(0.0431)
L21.GovernmentResponseIndex	-0.0759*
	(0.0435)
L28.GovernmentResponseIndex	-0.0164
	(0.0291)
Constant	18.40***
	(0.739)
Observations	40,878
Number of fips	2,736
R^2	0.661

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX E

Bucketed Regressions

Table E.1. Table of Regressions on Stratified Data

VARIABLES	(1) Democratic	(2) Moderate	(3) Republican	(4) Low Income	(5) Middle Income	(6) High Income
L14.NewCaseRate	0.815***	0.938***	0.933***	1.024***	0.846***	0.831***
	(0.00546)	(0.00224)	(0.00231)	(0.00235)	(0.00276)	(0.00291)
L14.weather	1.238***	1.319***	1.549***	1.093***	1.429***	1.253***
	(0.0737)	(0.0330)	(0.0383)	(0.0385)	(0.0413)	(0.0446)
L14.GovernmentResponseIndex	-0.0923***	-0.327***	-0.549***	-0.394***	-0.292***	-0.402***
	(0.0102)	(0.00626)	(0.00824)	(0.00752)	(0.00752)	(0.00844)
Constant	5.692***	17.01***	27.54***	20.01***	16.14***	22.93***
	(0.620)	(0.339)	(0.408)	(0.396)	(0.409)	(0.442)
Observations	20,033	126,999	145,564	100,227	89,547	98,811
R^2	0.543	0.617	0.601	0.693	0.547	0.506
Number of fips	167	1,093	1,503	930	785	981

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

REFERENCES

Avery, C., Bossert, W., Clark, A., Ellison, G., and Ellison, S. F. (2020). An economist's guide to epidemiology models of infectious disease. *Journal of Economic Perspectives*, 34(4):79–104.

FOX News Network (2021). Presidential election results. https://www.foxnews.com/elections/2020/general-results. Accessed: 2021-03-28.

Goldhaber, D., Imberman, S., Strunk, K., Hopkins, B., Brown, N., Harbatkin, E., and Kilbride, T. (2021). To what extent does in-person schooling contribute to the spread of covid-19? evidence from michigan and washington.

Hale, T., Webster, S., Petherick, A., Phillips, T., and Kira, B. (2020). Oxford covid-19 government response tracker, blavatnik school of government. Data use policy: Creative Commons Attribution CC BY standard.

Karaivanov, A., Lu, S. E., Shigeoka, H., Chen, C., and Pamplona, S. (2020). Face masks, public policies and slowing the spread of covid-19: Evidence from canada.

Mackenzie, D. (2021). *COVID-19: the pandemic that never should have happened, and how to stop the next one*. Little, Brown.

Murray, E. J. (2020). Epidemiology's time of need: Covid-19 calls for epidemic-related economics. *Journal of Economic Perspectives*, 34(4):105–120.

National Oceanic and Atmospheric Administration (2020). Global surface summary of the day. https://www.ncei.noaa.gov. Accessed: 2021-03-28.

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal: Promoting communications on statistics and Stata*, 13(1):65–76.

Simonov, A., Sacher, S., Dubé, J.-P., and Biswas, S. (2020). The persuasive effect of fox news: Non-compliance with social distancing during the covid-19 pandemic.

The New York Times (2021). Coronavirus (covid-19) data in the united states. https://github.com/nytimes/covid-19-data. Accessed: 2021-03-28.